

JOHN R. TAYLOR

---

LINGUISTIC  
CATEGORIZATION

Prototypes in Linguistic Theory



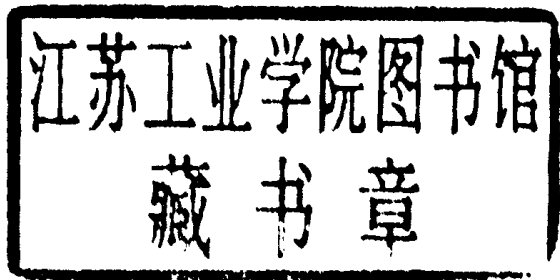
CLARENDON PRESS

OXFORD

# LINGUISTIC CATEGORIZATION

*Prototypes in Linguistic Theory*

JOHN R. TAYLOR



CLARENDON PRESS · OXFORD

1989

*Oxford University Press, Walton Street, Oxford OX2 6DP*

*Oxford New York Toronto  
Delhi Bombay Calcutta Madras Karachi  
Petaling Jaya Singapore Hong Kong Tokyo  
Nairobi Dar es Salaam Cape Town  
Melbourne Auckland  
and associated companies in  
Berlin Ibadan*

*Oxford is a trade mark of Oxford University Press*

*Published in the United States  
by Oxford University Press, New York*

© John R. Taylor 1989

*All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
electronic, mechanical, photocopying, recording, or otherwise, without  
the prior permission of Oxford University Press*

*British Library Cataloguing in Publication Data*

*Taylor, John R.*

*Linguistic categorization: prototypes in linguistic  
theory.*

*1. Cognitive linguistics*

*I. Title*

*410*

*ISBN 0-19-824838-5*

*Library of Congress Cataloging in Publication Data*

*Taylor, John R.*

*Linguistic categorization: prototypes in linguistic theory/John  
R. Taylor.*

*Bibliography. Includes index.*

*1. Categorization (Linguistics) 2. Linguistic analysis  
(Linguistics) 3. Grammar, Comparative and general—Grammatical  
categories. I. Title.*

*P128.C37T38 1989 410'.1—dc20 89-31949*

*ISBN 0-19-824838-5*

*Typeset by Joshua Associates Limited, Oxford*

*Printed in Great Britain by*

*Bookcraft (Bath) Ltd*

*Midsomer Norton, Avon*

*For  
Genia and Ary*

# Preface

THE title of this book is intentionally ambiguous. In one of its senses, 'linguistic categorization' refers to the process by which people, in using language, necessarily categorize the world around them. Whenever we use the word *dog* to refer to two different animals, or describe two different colour sensations by the same word, e.g. *red*, we are undertaking acts of categorization. Although different, the two entities are regarded in each case as the same.

Categorization is fundamental to all higher cognitive activity. Yet the seeing of sameness in difference raises deep philosophical problems. One extreme position, that of nominalism, claims that sameness is merely a matter of linguistic convention; the range of entities which may be called dogs, or the set of colours that may be described as red, have in reality nothing in common but their name. An equally extreme position is that of realism. Realism claims that categories like DOG and RED exist independently of language and its users, and that the words *dog* and *red* merely name these pre-existing categories. An alternative position is conceptualism. Conceptualism postulates that a word and the range of entities to which it may refer are mediated by a mental entity, i.e. a concept. It is in virtue of a speaker's knowledge of the concepts "dog" and "red", i.e. in virtue of his knowledge of the meanings of the words *dog* and *red*, that he is able to categorize different entities as dogs, different colours as red, and so on. Conceptualism may be given a nominalist or a realist orientation. On the one hand, we can claim that concepts merely reflect linguistic convention. The English speaker's concepts "red" and "dog" arise through his observation of how the words *red* and *dog* are conventionally used; once formed, the concepts will govern future linguistic performance. Alternatively, we might claim that concepts mirror really existing properties of the world. On this view, our concepts are not arbitrary creations of language, but constitute part of our understanding of what the world is 'really' like. This book will take a course which is intermediate between these two positions, yet strictly speaking consonant with neither. To the extent that a language is a conventionalized symbolic system, it is indeed the case that a language imposes a set of categories on its users. Conventionalized, however, does not necessarily imply arbitrary. The categories encoded in a

language are motivated, to varying degrees, by a number of factors—by actually existing discontinuities in the world, by the manner in which human beings interact, in a given culture, with the world, and by *general cognitive processes of concept formation*. It is precisely the dialectic of convention and motivation which gives rise to the fact that the categories encoded in one language do not always stand in a one-to-one correspondence with the categories of another language. Languages are indeed diverse in this respect; yet the diversity is not unconstrained.

In the first place, then, this book is about the meanings of linguistic forms, and the categorization of the world which a knowledge of these meanings entails. But language itself is also part of the world. In speaking of nouns, verbs, phonemes, and grammatical sentences, linguists are undertaking acts of categorization. The title of the book is to be understood in this second, reflexive sense. Just as a botanist is concerned with a botanical categorization of plants, so a linguist undertakes a linguistic categorization of linguistic objects. The second half of the book, in particular, will address the parallels between linguistic categorization in this second sense, and the categorization, through language, of the non-linguistic world. If, as will be argued, categories of linguistic objects are structured along the same lines as the more familiar semantic categories, then any insights we may gain into the categorization of the non-linguistic world may be profitably applied to the study of language structure itself.

The theoretical background to the study is a set of principles and assumptions that have recently come to be known as 'cognitive linguistics'. Cognitive linguistics does not (yet) constitute a theoretical paradigm which is able to rival, even less to displace, the (still) dominant generative-transformational approach. The main points of divergence are, however, clear. Whereas generativists regard knowledge of language as an autonomous component of the mind, independent, in principle, from other kinds of knowledge and from other cognitive skills, cognitivists posit an intimate, dialectic relationship between the structure and function of language on the one hand, and non-linguistic skills and knowledge on the other. Language, being at once both the creation of human cognition and an instrument in its service, is thus more likely than not to reflect, in its structure and functioning, more general cognitive abilities. One of the most important of these cognitive abilities is precisely the ability to categorize, i.e. to see similarity in diversity. A study of categorization processes

is thus likely to provide valuable insights into the meanings symbolized by linguistic forms. Furthermore, there is every reason to expect that the structural categories of language itself will be analogous, in many ways, to the categories which human beings perceive in the non-linguistic world around them.

The book owes its inception very largely to a suggestion from René Dirven. I am indebted to Professor Dirven, as well as to Maurice Aldridge, Brygida Rudzka-Ostyn, Dirk Geeraerts, and Savas Tsohatzidis for commenting on earlier versions of the manuscript. That the manuscript could be completed at all is due, in no small measure, to the constant encouragement, support, patience, and love, of my wife.

J. R. T.

# Typographical conventions

Linguistic forms are printed in italics: *dog*.

Meanings of linguistic forms, and glosses of foreign language forms, are given between double quotes: "dog".

Citations are marked by single quotes.

Names of categories are printed in small capitals: *DOG*.

Phonetic and semantic features are printed in small capitals enclosed in square brackets: [VOCALIC], [ANIMAL].

Semantic attributes are printed in normal type enclosed in square brackets: [ability to fly].

Phonemes, and phonemic transcriptions, are enclosed in slashes, phonetic symbols and phonetic transcriptions are enclosed in square brackets.

An asterisk \* indicates that a following linguistic expression is unacceptable, on either semantic or syntactic grounds. Expressions of questionable acceptability are preceded by a question mark.



# Contents

<i>Typographical Conventions</i>	xv
1. The Categorization of Colour	1
1.1 Why colour terms?	2
1.2 Arbitrariness	5
1.3 An alternative approach: focal colours	8
1.4 Autonomous linguistics vs. cognitive linguistics	16
2. The Classical Approach to Categorization	21
2.1 Aristotle	22
2.2 The classical approach in linguistics: phonology	24
2.3 The classical approach in semantics	29
3. Prototype Categories: I	38
3.1 Wittgenstein	38
3.2 Prototypes: an alternative to the classical theory	40
3.3 Basic level terms	46
3.4 Why prototype categories?	51
3.5 A note on fuzziness	54
3.6 Some applications	55
4. Prototype Categories: II	59
4.1 Prototypes	59
4.2 Prototypes and schemas	65
4.3 Folk categories and expert categories	68
4.4 Hedges	75
5. Linguistic and Encyclopaedic Knowledge	81
5.1 Domains and schemas	83
5.2 Frames and scripts	87
5.3 Perspectivization	90

5.4	Frames and scripts in language comprehension	91
5.5	Fake	92
5.6	Real	95
6.	Polysemy and Meaning Chains	99
6.1	Monosemous and polysemous categories	99
6.2	An illustration: Climb	105
6.3	Over	109
6.4	Some problems	116
7.	Category Extension: Metonymy and Metaphor	122
7.1	Metonymy	122
7.2	Metaphor	130
8.	Polysemous Categories in Morphology and Syntax	142
8.1	The diminutive	144
8.2	The past tense	149
8.3	A note on yes–no questions	154
9.	Polysemous Categories in Intonation	158
9.1	The problem of intonational meaning	158
9.2	The meanings of falling and rising tones	160
9.3	High key	168
10.	Grammatical Categories	173
10.1	Words, affixes, and clitics	175
10.2	Grammatical categories	183
10.3	The semantic basis of grammatical categories	190
11.	Syntactic Constructions as Prototype Categories	197
11.1	Constructions	198
11.2	The possessive genitive	202
11.3	The transitive construction	206
11.4	The transitive construction: more marginal members	210

	<i>Contents</i>	xiii
11.5	Metaphorical extension of syntactic constructions	215
11.6	A comparison with German	218
11.7	Concluding remarks	220
12.	Prototype Categories in Phonology	222
12.1	Phoneme categories	223
12.2	The gradience of phoneme categories	230
12.3	The syllable as a construction	234
13.	The Acquisition of Categories	239
13.1	Hypothesized acquisition routes	240
13.2	Grammatical categories	243
13.3	Conceptual development	247
13.4	Word meanings	252
	<i>References</i>	257
	<i>Index</i>	269

# 1

## The Categorization of Colour

As pointed out in the Preface, linguistics is concerned with categorization on two levels. In the first place, linguists need categories in order to describe the object of investigation. In this, linguists proceed just like practitioners of any other discipline. The noises that people make are categorized as linguistic or non-linguistic; linguistic noises are categorized as instances of a particular language, or of a dialect of a particular language; sentences are categorized as grammatical or ungrammatical; words are categorized as nouns and verbs; sound segments are classified as vowels or consonants, stops or fricatives, and so on.

But linguists are (or should be) concerned with categorization at another level. The things that linguists study—words, morphemes, syntactic structures, etc.—not only constitute categories in themselves, they also stand for categories. The phonetic form [red] can not only be categorized as, variously, an English word, an adjective, a syllable with a consonant–vowel–consonant structure; [red] also designates a range of physically and perceptually distinct properties of the real world (more precisely, a range of distinct visual sensations caused by the real-world properties), and assigns this range of properties to the category RED. The morphosyntactic category PAST TENSE (usually) categorizes states of affairs with respect to their anteriority to the moment of speaking; the preposition *on* (in one of its senses) categorizes the relationship between entities as one of contact, and so on.

Both in its methodology and in its substance, then, linguistics is intimately concerned with categorization. The point has been made by Labov (1973: 342): ‘If linguistics can be said to be any one thing it is the study of categories: that is, the study of how language translates meaning into sound through the categorization of reality into discrete units and sets of units.’ Questions like: Do categories have any basis in the real world, or are they merely constructs of the human mind? What is their internal structure? How are categories learnt? How do people go about assigning entities to a category? What kinds of

relationships exist amongst categories? must inevitably be of vital importance to linguists. Labov, in the passage just referred to, goes on to point out that categorization 'is such a fundamental and obvious part of linguistic activity that the properties of categories are normally assumed rather than studied'. In recent years, however, research in the cognitive sciences, especially cognitive psychology, has forced linguists to make explicit, and in some cases to rethink, their assumptions. In this first chapter, I will introduce some of the issues involved, taking as my cue the linguistic categorization of colour.

### 1.1 Why colour terms?

There are good reasons for starting with colour terms. In many respects colour terminology provides an ideal testing ground for theories of categorization. It is commonly asserted—by linguists, anthropologists, and others—that categories have neither a real-world nor a perceptual base. Reality is merely a diffuse continuum, and our categorization of it is ultimately a matter of convention, i.e. of learning. This view was expressed very clearly by the anthropologist, Edmund Leach:

I postulate that the physical and social environment of a young child is perceived as a continuum. It does not contain any intrinsically separate 'things'. The child, in due course, is taught to impose upon this environment a kind of discriminating grid which serves to distinguish the world as being composed of a large number of separate things, each labelled with a name. This world is a representation of our language categories, not vice versa. Because my mother tongue is English, it seems self evident that *bushes* and *trees* are different kinds of things. I would not think this unless I had been taught that it was the case. (Leach 1964: 34)

According to Leach, the categories that we perceive in the world are not objectively there. Rather, they have been forced upon us by the categories encoded in the language that we happen to have been brought up with. If categorization is language dependent, as Leach and many others suggest, it is only to be expected that different languages will encode different categorizations, none of them intrinsically any better founded, or more 'correct', than any other.

Intuitively, we would probably want to reject, on common-sense grounds, the idea that *all* categories are merely learnt cultural artefacts, the product of our language, with no objective basis in

reality. Surely, the world does contain discrete nameable entities, and in many cases there does seem to be a natural basis for grouping these entities into discrete categories. Tables are one kind of thing, distinct from chairs; elephants are another, and quite different from giraffes. These cases need not concern us at the moment. There is, though, one area of experience where the reality-as-a-continuum hypothesis would seem to hold, and this is colour. It has been estimated that the human eye can discriminate no fewer than 7.5 million just noticeable colour differences (cf. Brown and Lenneberg 1954). This vast range of visible colours constitutes a three-dimensional continuum, defined by the parameters of hue (the wavelength of reflected light), luminosity (the amount of light reflected), and saturation (freedom from dilution with white). Because each of these dimensions constitutes a smooth continuum, there is no physical basis for the demarcation of discrete colour categories. Yet people do recognize discrete categories. It follows—so the argument goes—that these categories are a product of a learning experience, more particularly, of language. This view is supported by the fact that languages differ very considerably, both with regard to the number of colour terms they possess, and with regard to the denotational range of these terms.

There are some well-known examples of non-correspondence of colour terms in different languages (see Lyons 1968: 56f.). Russian has no word for blue; *goluboy* “light, pale blue” and *sinii* “dark, bright blue” are different colours, not different shades of the same colour. *Brown* has no single equivalent in French; the range of colours denoted by *brown* would be described in French as *brun*, *marron*, even *jaune*. Welsh *glas* translates into English as *blue*, *green*, or even *grey*. Very often, it is not just an individual colour term which does not have an exact equivalent in another language. Rather, it is the set of colour terms as a whole which fails to correspond with that of another language. Bantu languages are on the whole rather poor in colour terms; Tsonga, for instance, has only seven basic colour terms.<sup>1</sup> These, with their approximate range of English equivalents, are as follows:

- (1) ntima: black
- rikuma: grey
- basa: white, beige

<sup>1</sup> The notion of basic colour terms will be elaborated later, in s. 1.3. In addition to their basic colour terms, both Tsonga and Classical Latin (to be discussed below) have a large number of non-basic terms which denote quite precisely the colours characteristically associated with particular kinds of object.

tshwuka: red, pink, purple  
 xitshopana: yellow, orange  
 rihlaza: green, blue  
 ribungu: dark brown, dull yellowish-brown

Tsonga divides the black–grey–white dimension in essentially the same way as English. However, only three categories are recognized in the hue dimension (*tshwuka*, *xitshopana*, *rihlaza*), whereas English has at least six (*purple*, *red*, *orange*, *yellow*, *green*, *blue*). *Ribungu*, on the other hand, is a special word for colours of low luminosity in the yellow–orange–brown region. Neither do we need go to non-European languages to find cases of extensive non-correspondence with English terms. Older European languages typically exhibit rather restricted colour vocabularies, which contrast strikingly with the modern English system. Consider the colour terms in Classical Latin (André 1949):

- (2) *albus*: white  
*candidus*: brilliant, bright white  
*ater*: black  
*niger*: shiny black  
*ruber*: red, pink, purple, orange, some shades of brown  
*flavus*: yellow, light brown, golden red  
*viridis*: green  
*caeruleus*: blue

We find here, as in Tsonga, a rather restricted range of terms for the hue dimension. On the other hand, Latin made a distinction, lacking in English, between blacks and whites of high and low luminosity.

Linguists have not been slow to recognize the theoretical significance of colour terminology. Consider the following passage from Bloomfield's classic volume *Language*:

Physicists view the color-spectrum as a continuous scale of light-waves of different lengths, ranging from 40 to 72 hundred-thousandths of a millimetre, but languages mark off different parts of this scale quite arbitrarily and without precise limits, in the meanings of such color-names as *violet*, *blue*, *green*, *yellow*, *orange*, *red*, and the color-names of different languages do not embrace the same gradations. (Bloomfield 1933: 140)

This passage by Bloomfield could have been the model for Gleason's treatment of the same topic in his once very influential *Introduction to Descriptive Linguistics*:

Consider a rainbow or a spectrum from a prism. There is a continuous gradation of color from one end to the other. That is, at any point there is only a small difference in the colors immediately adjacent at either side. Yet an American describing it will list the hues as *red, orange, yellow, green, blue, purple*, or something of the kind. The continuous gradation of color which exists in nature is represented in language by a series of discrete categories. . . . There is nothing inherent either in the spectrum or the human perception of it which would compel its division in this way. The specific method of division is part of the structure of English. (Gleason 1955: 4).

Other statements in the same vein could be quoted from other scholars. Indeed, many textbooks and surveys of linguistic theory (the present work is no exception!) have an obligatory paragraph, even a whole section or chapter, devoted to colour.

I would like to draw attention to one particularly important detail in the passage from Bloomfield, namely the assertion that colour categorization is *arbitrary*. Gleason, a few pages after the above quotation, makes the same point. What is more, Gleason puts his discussion of colour in the very first chapter of his textbook, as if to suggest that the arbitrariness of colour terms is paradigmatic for the arbitrariness of language as a whole. The arbitrariness of colour terms follows from the facts outlined above, namely the physical continuity of the colour space, and the human ability to make an incredibly large number of perceptual discriminations. There are, no doubt, other areas of experience which, like colour, constitute a smooth continuum: length, height, temperature, speed, perhaps even emotions like love, hatred, anger. Human beings can also make a large number of perceptual discriminations in these domains (but presumably nothing like the alleged 7.5 million colour discriminations). Languages are typically rather poor in their categorization of these domains. For length, English has only two terms, *long* and *short*. Colour, with its rich and language-specific terminology, is indeed an ideal hunting ground for anyone wishing to argue the arbitrariness of linguistic categories.

## 1.2 Arbitrariness

Arbitrariness, as I have used the term in the preceding paragraph, has been a *fundamental concept* in twentieth-century linguistics. Its status as a quasi-technical term goes back to Saussure, who, in his *Cours de*



*linguistique générale* (1916) proclaimed as a first principle of linguistic description that ‘the linguistic sign is arbitrary’: ‘le signe linguistique est arbitraire’ (Saussure 1964: 100).

The linguistic sign, for Saussure, is the association of a form (or signifier) with a meaning (or signified). There are two respects in which the linguistic sign is arbitrary (see Culler 1976: 19 ff.). In the first place, the association of a particular form with a particular meaning is arbitrary. There is no reason (other than convention) why the phonetic form [red] should be associated with the meaning “red” in English; any other phonetic form, provided it was accepted by the generality of English speakers, would do equally well. It is therefore to be expected that different languages will associate quite different phonetic forms with a particular meaning; were the relationship not arbitrary, words with the same meaning in different languages would all have a recognizably similar form. With this characterization of arbitrariness, few would disagree.<sup>2</sup> But there is another, more subtle aspect to arbitrariness, as Saussure conceived it. This is that the signified itself—the meaning associated with a linguistic form—is arbitrary. Saussure vigorously denied that there are pre-existing meanings (such as “red”, “orange”, etc.), which are there, independent of language, waiting to be named. The lexicon of a language is not simply a nomenclature for some universally valid inventory of concepts. There is no reason, therefore, why any portion of the colour space should have a privileged status for categorization in the colour vocabulary of a language; indeed, strictly speaking, there is no reason why colour should be lexicalized at all. We return, then, to the theme of Section 1.1. Reality is a diffuse continuum, and our categorization of it is merely an artefact of culture and language.

The arbitrariness of the linguistic sign is closely linked to another Saussurian principle, namely the notion of language as a self-contained, autonomous system. ‘La langue’, according to Saussure, ‘est un système dont tous les termes sont solidaires et où la valeur de l’un ne résulte que de la présence simultanée des autres’ (1964: 159). The meaning of a linguistic sign is not a fixed property of the linguistic sign considered in and of itself; rather, meaning is a function of the value of the sign within the sign system which constitutes a language.

<sup>2</sup> The doctrine of the arbitrariness of the signifier–signified relationship disregards, of course, the relatively rare phenomena of onomatopoeia and sound symbolism. It is worth mentioning that Rhodes and Lawler (1981) have recently suggested that the phonetic motivation of the signifier might be much more extensive than is traditionally believed.